MULTI-MODAL WORKLOAD IMPACTS ON BATTLEFIELD SITUATION AWARENESS AND PRIMARY TASK PERFORMANCE

A Thesis
Presented to
The Academic Faculty

by

Brandon Alexander Thomas

In Partial Fulfillment
of the Requirements for the Degree
of Master's of Science in the
School of Psychology

Georgia Institute of Technology
December 2016

MULTI-MODAL WORKLOAD IMPACTS ON BATTLEFIELD SITUATION AWARENESS AND PRIMARY TASK PERFORMANCE

Approved by:

Dr. Bruce Walker, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Dr. Rick Thomas
School of Psychology
Georgia Institute of Technology

Date Approved: July 17th, 2016

To the students of the Georgia Institute of Technology, thank you for you random smiles along the way. Bruce Walker, Thom Gable as well as the rest of the Sonlab and members of the School of Psychology own my everlasting gratitude for helping me achieve my goals at becoming an Instructor at the United States Military Academy. Your efforts will be forever entwined with the future leadership of America's Army. To BGSA, thank you for being family away from home. Last but not least, thank to my beautiful wife Christina Thomas who assisted with countless proof reads and to Elizabeth our daughter that reminded me there is always time for a hug and raucous laughter.

ACKNOWLEDGEMENTS

I wish to thank in addition to my previous dedication, my mother and father for encouragement to stay the course and not to sweat the small stuff. In addition, I would like to thank Jackie Cox, Tia Jackson Truitt and Felicia Benton-Johnson for their constant encouragement and open ears. Their smiles made my days fly by while at Georgia Tech and while working with them on special projects for the Center for Engineering Education and Diversity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The U.S. Military is downsizing and streamlining its forces both in personnel and equipment; yet, the Department of Defense has promised to do so while leveraging technological advantages through proliferation of technology to fight and win America's wars. Commanders today must execute an enormous task load in an overwhelming environment of current tasks and associated technological requirements. This study examined the simple question, how much is too much technology? This was done through operationalizing visual, auditory, cognitive, and psychomotor workloads on battlefield situation awareness. Through the use of a primary task, reporting task, and N-back task, participants baselined their performance and then experienced workload level manipulations within the tasks. Ultimately, the study identified that there is a threshold for performance and situation awareness maintenance in complex workload environments, such as those a military commander currently finds themselves, due to the taxation on executive functions regardless of the VACP workload type. The future will lead to implementation of more technology and subsequent VACP workload requirements. In exploring the effects of varied workload combinations and their direct impact on performance and situation awareness, this research found that a balance must be achieved in human-centered systems.

CHAPTER 1

MILITARY TECHNOLOGICAL PROGRESSION

The U.S. Military is downsizing and streamlining its forces both in personnel and equipment; yet, the Department of Defense has promised to "sustain priority investments in science, technology, research, and development… [by] actively seeking innovative approaches to how we fight…and how we leverage our asymmetric strengths and technological advantages" (Department of Defense: 2014 Quadrennial Defense Review, 2014, D.S. Hagel's letter). In plain English this means that with fewer personnel, the military will be forced to rely on technology to fill in gaps left by this decrease. To that end, there are several programs designed around the operation of modular vehicle battle management systems. In 2004, the penultimate technological edge at the tactical level in the Army was a digital reporting and battle management tracking system called FBCB2 or Force XX1 Battle Command Brigade and Below. "The system provides mobile, near real-time battle management of multiple assets and aids in situation awareness" (Durlach, 2004). Studies concerning attention during operation of FBCB2 user interface systems abound and have even endured to this day in the civilian realm of cellular phone technologies and in vehicle infotainment systems (Davies & Beeharee, 2012); however, very little work has been done concerning situation awareness maintenance while performing multiple tasks that require use of several senses.

A brief examination of the military ground commander's tasks range in scope, (Headquarters, Department of the Army, 2012). Going into detail of each task is beyond the scope of this document; however, commanders must execute an enormous task load and there is already an overwhelming nature of current tasks and associated technological requirements.

- Current and proposed military technology systems also range in function and scope; however, the magnitude of anticipated reliance on technological systems that utilize human Visual, Auditory, Cognitive and Psychomotor (VACP) capabilities, specifically, information display and auditory interfaces are profound. The Department of Defense's *Unmanned Systems Integrated Roadmap (USIR) FY2014- 2038*, outlines in detail the projected glide-path for joint research and development of military systems. Figures 1 and 2 give a brief projection of unmanned aircraft and ground vehicle systems from the years 2013 to 2038.

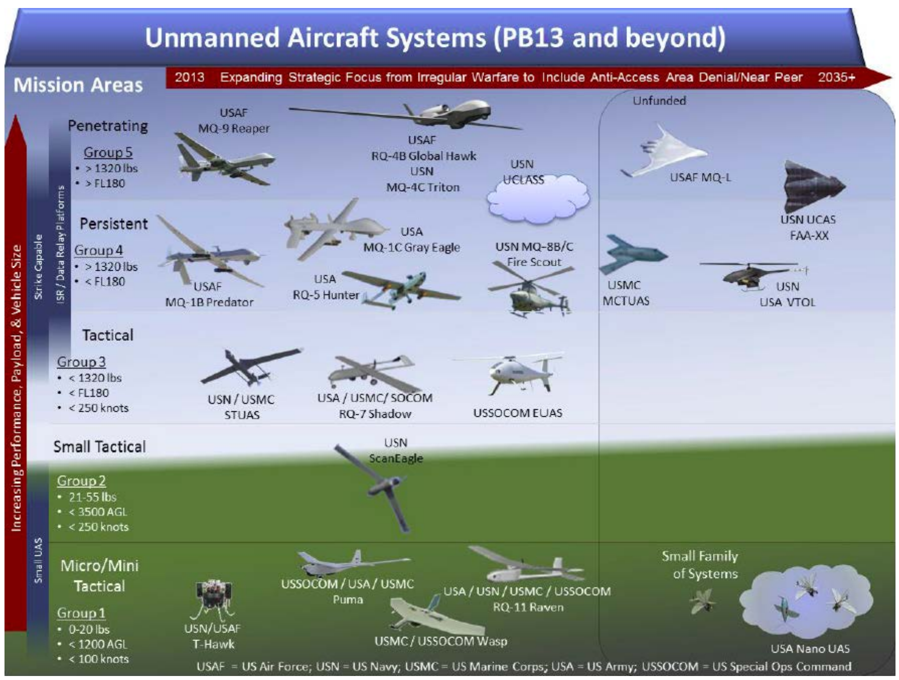

Figure 1. Unmanned Aircraft Systems Fiscal Year 2013-2038

The proposed technology and advancements of the future will undoubtedly increasingly rely on human-system interaction to the potential detriment of user performance based on current Military Decision Making procedures (Headquarters, Department of the Army, 2012). A cursory examination of just one of these systems, a route probe robot being operated from a moving vehicle, broaches the concept of alternate perspective in visual displacement, as well as, conflicts with sensory orientation and movement that contrast with the display and controls. Human performance has been shown to decrease with high workloads (Scribner, Wiley, Harper, & Kelley, 2007, Robert & Hockey, 1997, Wickens, 2008, Lim, Wu, Wang, Detre, Dinges, & Rao, 2010). The impact of such stimuli on the human operator's performance with such a heavy workload calls into question how humans can operate under such high workloads and still perform their tasks.

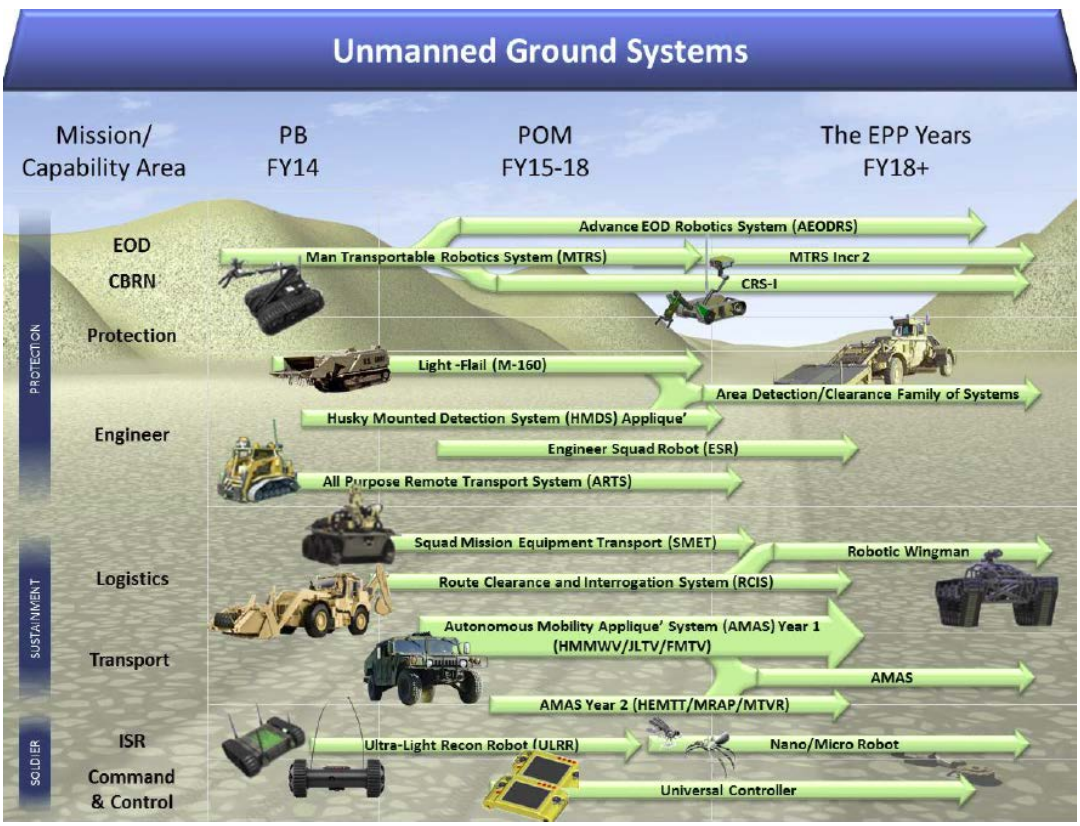

Figure 2. Unmanned Ground Systems by Mission / Capability FY14-20

CHAPTER 2

MENTAL WORKLOAD THEORY

While research abounds in the realm of mental workload theory, there are no universally accepted theories or even definitions. Many of theories stem from information-processing models in which the human system operates similar to a computer. In both systems information is input, encoded, and given meaning based on schemas or previously stored information and an output or task is executed. With regards to specific research done by the military, the Army Research Laboratories have been at the forefront of investigating mental workload. They conducted a review of mental workload theories, choosing to model their work on research by Wickens's (1991) multiple resource theory (MRT). MRT states, that there are different pools of resources humans can draw on in parallel when discussing the importance of attention resources. Furthermore, MRT explains a lapse in performance as a result of approaching or reaching a threshold in one or more of those resource channels, because "the human has a limited capacity for processing information" (Mitchell, 2000).

The best framework for identifying and subsequently quantifying workload attributes came from the ARL model that simply defines the resource channels as Visual, Auditory, Cognitive and Psychomotor (VACP) workloads and further describes a model with which to theoretically predict workload capacities through simulations. Their review was thorough; however, the evaluation system had not been evaluated on real-world systems in an experimental fashion to verify the predictive model (Mitchell, 2000). Collectively, their theoretical work and modeling allows for system designers to predict which tasks will interfere with others when more than one task is performed simultaneously and subsequently, which task will suffer performance decrements. For example, if the visual system is bombarded by sensory information for processing by our "brain computers", flashing alerts or changes to the visual information may overload the optimal state of the system and a message may go unattended or processed. The same is true for all senses, overwhelming the auditory channel will cause verbal conversations of information added to the system will prevent encoding and processing of some input components. Additional descriptions of workload theories and associated models can be found in Huey and Wickens (1993), Damos (1991), Kramer (1991), Liao & Moray (1993), and Wickens (2004). The following study sought to parse the workload of a vehicle commander into manipulable and quantifiable attributes. Based on MRT framework, this study isolated components of the VACP model and paired them with three tasks that allowed for manipulation of the VACP loads on each resource channel through a visual and psychomotor Ball Drop Task, an auditory and cognitive N-Back Task, and finally, a visual battlefield awareness interface comprised of an auditory or visual stimulus toggle.

CHAPTER 3

SITUATION AWARENESS

The components from the MRT models, and in this case, the VACP workloads, become invariably linked with situation awareness as they require attention to "input" and mental workload to "encode" within the selected models. Situation awareness is similar to workload theory because a definition is ill defined. Situation awareness pertains to the ability to maintain a current mental model of events and information. Here, situation awareness will be defined according to Endsley's definition that it does "not encompass[ing] all of a person's knowledge… [but] only [to] that portion pertaining to the state of a dynamic environment" (1995, Pg. 36). Additionally, Endsley's works in situation awareness refer to the three levels of situation awareness (1995) – the first being perception of the elements in the environment, the second is the comprehension of the current situation and the third being projection of future status in the dynamic situation. These attributes of situation awareness are important because they occur in the closed "processing unit" of the human mind and require some level of mental workload to achieve the end result of knowledge about a dynamic environment. Because these levels of situation awareness have potential to load working memory, Endsley's theory assumes that individuals will "recognize key features in the environment – critical cues – that will map to key features in the model…[thereby providing] for the higher levels of SA (comprehension and projection)" (1995, pg. 44).

Analysis of SA can come in many forms, with each having its strengths and weaknesses. One such common method used to measure SA is Endsley's Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995b; 2012).

> SAGAT involves asking operators queries about the systems they are controlling (e.g., what is the current altitude of your aircraft? How close will your aircraft be to the nearest weather cell in the next 5 minutes?). Most SAGAT queries require a "yes/no" or multiple choice response format, although some can involve asking the operator to reproduce the state of specific events. More specifically, SAGAT involves freezing simulations at random intervals and blanking displays prior to the presentation of probe questions. (Vu & Chiappe, 2015)

Additional SA evaluation tools exist such as the Situation Present Awareness Model (SPAM), which can remove the aspects of memory and recall from analyzing situation awareness making it a situation assessment (Durso, Hackworth, Truitt, Crutchfield, & Nikolic, 1999). This method allows for isolation of memory variables associated with other methods such as SAGAT. The SPAM process allows researchers to measure SA as a process and not a product because it involves measuring accuracy and response times of participants' answers after a ready prompt is given. In this method, they have full access to the information through the system interface; hence, eradicating induced error from the memory variable. It should now be clear that workloads and SA share common cognitive components and there is a large potential for interaction with each other that can greatly effect performance, especially in future military systems.

CHAPTER 4

MILITARY APPLICATIONS OF CURRENT THEORIES

Workload and SA in Current In-Vehicle Military Systems

Even though we know what the military wants and potentially what the projected technology will require, we can examine the current systems to see that there already exists a plethora of systems and in-vehicle technologies that require a vast amount of attentional resources. Even a brief list of in-vehicle ground systems in use today or projected for use in the next 10 years is beyond the scope of this document. It is sufficient instead, to summarize some current and future systems in an effort to exhibit general technology implementation trends. The in-vehicle battle management system, FBCB2 mentioned previously, allows for – navigation, reporting / communication, and intelligence updates all in near real-time through satellite communication (SATCOM). Technology such as SATCOM, is used in scenarios where the current radio and communication technologies are insufficient. Most cases involve a long distance between two elements whether those elements are systems or people and regardless of their location on the land, sea or in the air. The distances we are referencing will remain unspecified; however, they could potentially range from 5 to 500 kilometers depending on terrain and other factors.

Regardless of the method of communication, the military still has current in-vehicle technology that allows them to operate systems remotely though interfaces. The Common Remotely Operated Weapon System (CROW), provides a plethora of information to include ballistic solutions, weapon feedback data via the Remote Weapon System (RWS) display, and several visual detection modes (e.g. thermal imaging). While the system is currently used by the gunner on vehicle platforms, not by the driver or co-driver, it has the potential for any crew position to operate it and is less technologically advanced in wheeled systems as compared to the M1A2 SEP 3 Abrams tank commander weapons system. Due to the modular nature of the aforementioned systems and the fact that Soldiers physically operate these systems through visual displays and control apparatus, many studies have been conducted involving the interactions between humans and the visual display systems in military contexts from healthcare to battlefield tracking (Zhang et al.; Sweeny, 2008). The one common thread between the majorities of current systems in the military is that they place a large visual, auditory, cognitive and psychomotor load on the operator because of monitoring, tracking, targeting, information dissemination, and even executing defensive and offensive operations.

Even though there is a heavy reliance on visually based systems in military vehicles, visual systems are not the only information technology present and the other requirements utilize additional modalities of human resources. Situation updates, for example, are continuously pushed from the top-down via auditory alerts on digital systems or radio communications. There is also a bottom-up load from the senior-subordinate military hierarchy that requires continuous updates as events and situations change via both visual and auditory modalities. Passive monitoring systems also exist in the form of threat contact identification with pertinent distance and direction data via the “Boomerang” or sonic warfare auditory alert systems. All of these additional auditory based systems directly point to the possibility that performance in a system of systems with high auditory, psychomotor, visual, and cognitive demands potentially overload the military individual when real-time decision making is required for survival.

Military Applications of Auditory Cues

To quell an argument that not every modality is overloaded requires an examination of the military's use of auditory cues as alleviation and their impact on the system of systems. Assuming that there is a large load on the user in the visual modality and not as much in the auditory modality, several studies have been conducted in the auditory cue line of research on enhanced auditory cues (Jeon, Davison, Nees, Wilson, & Walker, 2009; Jeon & Walker, 2009; Landsdown, Brook-Carter, & Kersloot, 2004; Gable, Walker, Moses, & Chitloor, 2013, Jeon, Walker & Gable, 2015). Auditory demands in military crews have had cursory studies conducted such as in Lenne, Hoggan, Fidock, Stuart, and Aidman (2014) where they found individuals prioritized and protected their primary visual / driving tasks to the detriment of their secondary auditory task. Basically they found that people stopped attending to the auditory task in favor of the primary or driving task and as the secondary auditory task condition increased, the subjects' performance on that task decreased. Presumably, this occurred due to two main factors – perceived risk associated with the primary search and driving tasks (due to safety concerns) and contradictory auditory / visual cognitive loading on unrelated tasks. A closely related civilian study by Engström, Johansson, and Östlund (2005) found that the visual and cognitive loads associated with task performance while driving in the real-world have "radically different effects on driving performance" (p. 117). Their study found that in visual tasks, time sharing between the task and road conditions or external environment, cause a shift in strategies "where the driver strives to maintain acceptable lane keeping performance by means of speed reduction and/or large steering corrections. By contrasts, cognitive load leads to gaze concentration towards the road center associated with increased lane keeping performance.” (Engström, Johansson, & Östlund, 2005, p. 117). That study is particularly relevant due to the high level of external validity. They tested three groups in varied simulated and real-world driving conditions to compare the effects and treatments as validly as possible, while attempting to make a correlation between simulators and real-world driving. The only significant difference found between conditions was that the physiological workload and steering activity was higher in the field. They attributed that to the higher actualized risk while driving in real traffic and as mentioned above, was apparent as an explanation of task prioritization in Lenne’s study.

While the demands of visual load, cognitive load, and auditory loads have been assessed to some degree, the interaction between those loads on individual performance have never been addressed to identify the impacts in-vehicle technologies have. Where civilian and military in-vehicle technologies differ, is the concept of user-initiated demands versus forced demands associated with real-time information and situation awareness priorities. A handful of research has been conducted on in-vehicle information systems and driver effects in the aforementioned domains in both simulated driving and real-world driving on multiple environment types (urban, rural, etc.) (Santos, Merat, Mouta, Brookhuis, & Waard, 2005; Strayer, Drews, & Johnston, 2003). While this study does not address the driving aspect, we have replaced driving with another measurable, performance based task of a Ball Drop “game” previously used to replicate driving (see Jeon, Davison, Nees, Wilson, & Walker, 2009, p. 93 for a detailed description of the task comparison).

Study Execution

What all of the government reviews fail to address are the projected VACP workloads and situation awareness demands these new technologies will place on the human operators.

This study is intended to further the line of research started by Lenne (2014); specifically, addressing the impact of multi-modal workloads on performance and situation awareness maintenance in a simulated battlefield awareness system. This study measured variations of task loads and type of load based on the Verbal Auditory Cognitive Psychomotor (VACP) model in depth on a ball drop task that took the place for the amount of resource load experienced by a vehicle commander during a combat mission. In particular, the study executed a focused expansion to Lenne (2014) with a number of extended research questions and parameters that primarily focus on the vehicle commander and their workloads. The new extension of research aimed to investigate the effect of manipulating workload type and level on the ability to maintain battlefield situation awareness on our interface that emulates current systems. The method and design was such that it sought to take advantage of a combination of empirical techniques most often used by workload researchers through the use of reaction times (physiological measures) and subjective measures such as the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988).

CHAPTER 5

METHOD

Participants

The sample was composed of 40 undergraduate students from a large university in the southeastern United States. All participants were self-selected through a research database and received partial class credit for participating in the study. Participants were required to have normal or corrected to normal vision, full color vision, normal or corrected to normal hearing, and efficient mobility in order to accomplish the three tasks of this study. Participants completed consent forms and reviewed the outline of the study as part of their in-brief. They also underwent a training period on the systems to get them beyond a novice level of experience.

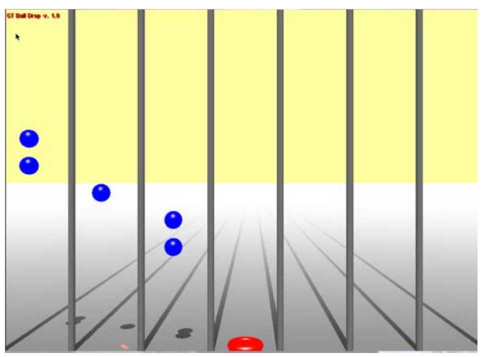

Figure 3. Ball Drop Screen as presented to user.

Apparatus

Primary task. The set-up consisted of a single workstation. The primary task was to perform a ball drop task on a 15-inch dell computer monitor. (See the ball drop task and protocol in Jeon et. al., 2009.) The ball drop task replicated the psychomotor requirements of a vehicle commander having to enter keyboard input to manipulate a system and accomplish a task within their vehicle workspace. The task also replicated the attention and cognitive load associated with in-vehicle navigation. Balls randomly “drop” towards the participant on a horizontal plane within six lanes (Figure 3). The participant had to move a receptacle to “catch” the ball using the left and right arrow keys on the keyboard.

This task consisted of two levels of manipulation concerning the rate of ball drop so that each participant performed at their established baseline for each level. The baseline was set through an established technique of running a pre-trial series of ball drops to set performance levels within the system to each participant. This technique ensured that the 90% and 60% performance level was standardized across each participants’ inherent capability.

Secondary task. The secondary task was for the participant to maintain situation awareness of a battlefield awareness emulator with a varied refresh rate. This task emulates a version of a military battlefield awareness and communication system previously indicated as the FBCB2. The interface was also displayed on a 15-inch Dell monitor that was coded to toggle between two SPAM presentation modes. The toggle function was established as part of the protocol and was only varied between participants for their designated group of Auditory or Verbal SPAM. In the Auditory SPAM condition, participants were presented with an auditory “Ready” cue from the higher command via the system voice speech file, replicating a request from the operations center that verbally says “Ready”, followed by an Auditory SPAM question after they indicated they were ready for the question (in accordance to SPAM protocols previously discussed). This condition can be increased in complexity in further studies by switching verbal command with text input and visual command with verbal input.

Their answers were recorded via a JVC portable video recording device to be later encoded for analysis. In the Verbal SPAM condition, participants were presented with a visual text button with the word “Ready” displayed under the system status buttons on the lower left portion of the screen. The participant had to use a standard USB connected mouse to click the “Ready” button. After the ready button was clicked, a text box at the top-center of the display would display a SPAM question and required text input as an answer via a standard American keyboard. Simply pressing the “Enter” key submitted their answer into the data output. All “ready” prompts were presented at varied times across each participant’s four trials in order to prevent anticipation of system situational cues; however, each trial across participants (Trials 1 through 4) presented the “ready” prompt at the exact same time (e.g., all Trial 1 “ready” prompts were scheduled across the nine minute trial run time at 35 seconds, 1 minute 18 seconds, etc. with Trial 2 prompts beginning at 22 seconds, 1 minute 30 seconds, etc.).

Cognitive load. Cognitive load was assessed by performance measures through the ball drop task. Cognitive load was also measured by an N-Back task at either the 0-back or 1-back level (Herff, Heger, Fortmann, Hennrich, Putze, & Schultz, 2013). After each scenario, we administered the NASA Task Load Index (Hart & Staveland, 1988). Wickens’s chapter on mental workload indicates that while quantitative methods of measuring cognitive workload exists, the qualitative, specifically the subjective measures are most commonly used. The industry standard is called the NASA Task Load Index (NASA-TLX) and measures six domains of demand - mental demand, physical demand, temporal demand, performance, effort, and frustration (Wickens, Hollands, Banbury, & Parasauraman, 2015). For an in depth discussion of NASA-TLX and its uses in performance studies see Landsdown et al. (2004). Subjective measures are also important because in some cases, participants devise management strategies to cope with higher workloads so actual performance measures may not indicate strain in the sense that errors and reaction times increase significantly (Mitchell, 2000).

Other measures. Participants also completed a demographic survey including gender, age, years of experience using navigation devices such as Google Maps, as well as any strategies, if any, they used to accomplish their various tasks.

Procedure

The study was presented in four experimental trials broken down into three tiers of difficulty, representing each condition paired with various levels of the other conditions so that each – 0-Back and 1-Back and Low WL level (90% performance)- and High WL level (60% performance)-level of the Ball Drop – were fully explored (see Table 1 for block mapping). In addition, we randomly assigned participants to either the Auditory presentation/verbal response SPAM group or Visual presentation/Manual text entry SPAM group (See Figure 4).

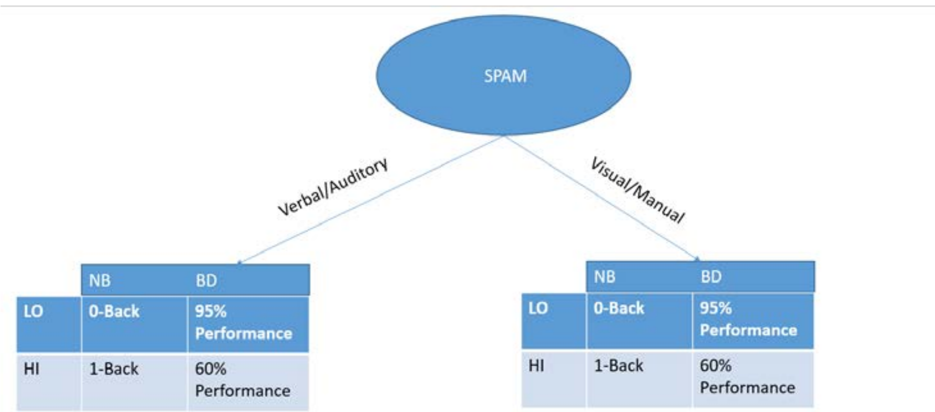

Figure 4. Experimental Design

After the participants completed a tier one WL level in Trial 1 consisting of a low WL condition with 90% performance on the ball drop task and a low WL condition with a 0-Back, they moved to tier two consisting of mixed ball drop and N-Back levels in Trial 2 and 3, ending with tier three (in terms of difficulty) in Trial 4 which consisted of high WL condition of 60% performance on the ball drop task and a high WL condition of a 1-Back recall level. After Trial 4 they completed the demographic questionnaire and received a debriefing on their participation.

Hypotheses

H 1-1. It was hypothesized that the results would show that participants would have the best ball drop performance and SA when there were no competing modalities in the first trial, meaning that Ball drop performance would be better in the Auditory SPAM, than in the Visual SPAM.

H 1-2A**.** Between the groups, the N-back performance was expected to show better performance in the Visual SPAM group due to available auditory resources, but should show a decline in their BD performance.

H 1-2B. There was also an expected significant decrease in performance in all measures except the N-back task when the Visual SPAM group enters into Trial 2 and Trial 3.

H 2-1. Furthermore, the SPAM task was expected to suffer the most degradation in Trial 4 tasks in the auditory SPAM group. In all cases, there should be significant evidence of a decrease in performance of a task when that task takes place simultaneously with another task in a like modality. Those Specific hypotheses were that:

H 2-1A. In Trial 4 for the Visual SPAM group should result in lower BD and N-Back performance (BD more so than N-back).

H 2-1B. In Trial 4 for the Auditory SPAM should result in lower BD and N-back performance (N-Back more so than BD).

CHAPTER 6

RESULTS

The manipulation of changing the level of Ball Drop performance from low WL (90% performance) to high WL (60% performance) concerning percentages of balls caught was successful in affecting the percentage means of balls caught. The high WL condition of BD performance level at 60% of balls caught led to a significantly lower performance in the BD task (BD percentage of balls caught means = 74.57 and 52.00% with standard deviations = 1.69 and 1.87% for the low WL and the high WL BD performance, respectively), $F(1,38) = 212.818$, $p < 0.01$. The manipulation of changing the level of N-back from low WL of 0-back to a higher WL of 1-back was successful in affecting both performance on the N-back task. The more difficult N-back recall of 1-back led to significantly lower performance in the N-back task (N-Back percentage correct means = 91.10 and 85.50% for the 0-back and 1-back, respectively),

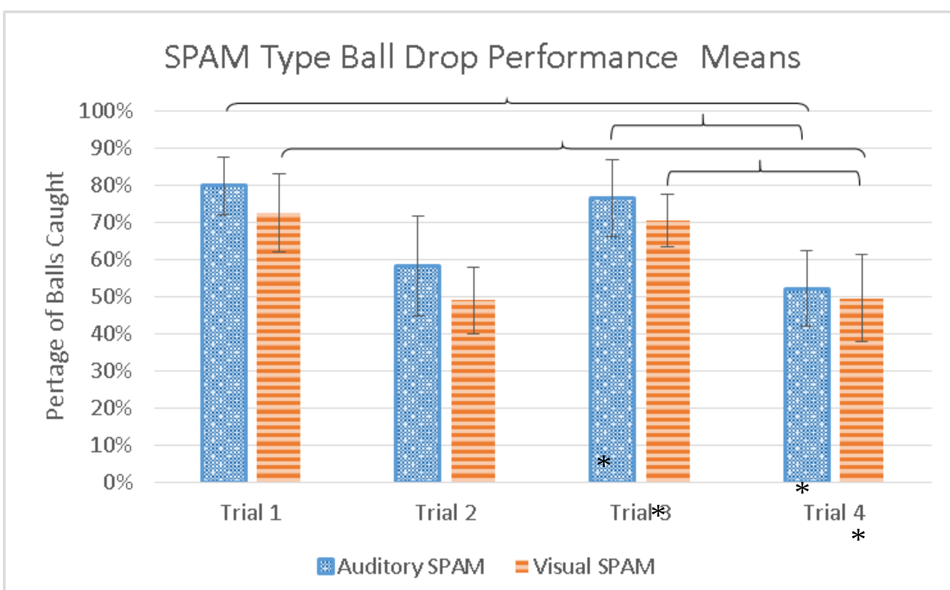

Figure 5. SPAM Type Ball Drop Performance Means

$F(1,38) = 8.489$ $p < 0.01$. For a complete breakdown of means, by group and trial, please see Table 1 in Appendix A.)

H 1-1 Results. The first hypothesis stated that the results should show that participants would have the best ball drop performance and SA when there are no competing modalities in the first trial (e.g. Ball drop performance will be better in the Auditory SPAM, than in the Visual SPAM). As Figure 5 indicates, a review of the analysis for the BD comparisons between the Auditory and Visual SPAM groups at each trial level indicated that in Trial 1 (left most bar pair in Figure 5), the Auditory SPAM group performed better than the Visual SPAM group (Mean BD Performance percentages = 79.75 and 72.54 with SD = 7.64 and 10.62 for the Auditory and Visual SPAM groups, respectively). The independent samples t-test of these means shows that the Auditory SPAM group did catch more balls when compared to the Visual SPAM group at the 90% performance level when the N-back task was held at 0-back for both groups ($t(40) =2.543$, $p<.05$). The second pair of bars shows that by decreasing the BD performance level to 60% performance and holding the N-back at the low WL condition of 0-back as in Trial 1, the Auditory SPAM group still performed better than the Visual SPAM group ($t(40) = 2.60$, $p<.05$). In Trial 3 (bar pair three in Figure 5), the BD performance level was calibrated back to the low WL condition of 90% performance; however, the N-back was increased to the high WL condition of the 1-back level. The Auditory SPAM group, again outperformed the Visual SPAM group in catching more balls in the third trial ($t(40)= 2.18$, $p <.05$). The MANOVA showed that the Auditory SPAM group outperformed the Visual SPAM group in number of balls caught on the Ball Drop task, $F(1,38) = 10.379$, $p < .01$.

H 1-2A Results. Between the groups, the N-back scores were expected to show better performance in the Visual SPAM group due to available auditory resources, but should also show a decline in their BD performance. The middle two bar pairs in Figure 6 show the change in performance levels across Trial 2 (Mean BD Performance percentages = 58.11 and 48.89 with SD = 13.43 and 8.84 for the Auditory and Visual SPAM groups, respectively) and Trial 3 (Mean BD Performance percentages = 76.47 and 70.45 with SD = 10.34 and 7.02 for the Auditory and Visual SPAM groups, respectively). In both Trial 2 and Trial 3, the Auditory SPAM group numerically outperformed the Visual SPAM group in the Ball Drop task; Trial 2, $t(40)=2.599$, $p<.05$ and Trial 3, $t(40)=2.183$, $p<.05$.

When comparing BD performances in Trial 4 (right most bar pair in Figure 5) between the Auditory and Visual group, the means dropped to a level that did not significantly indicate a difference between the groups and that the participants potentially reached a performance ceiling based on the 60% BD performance and 1-back recall (Mean BD Performance percentages = 52.05 and 49.41 with SD = 10.21 and 11.70 for

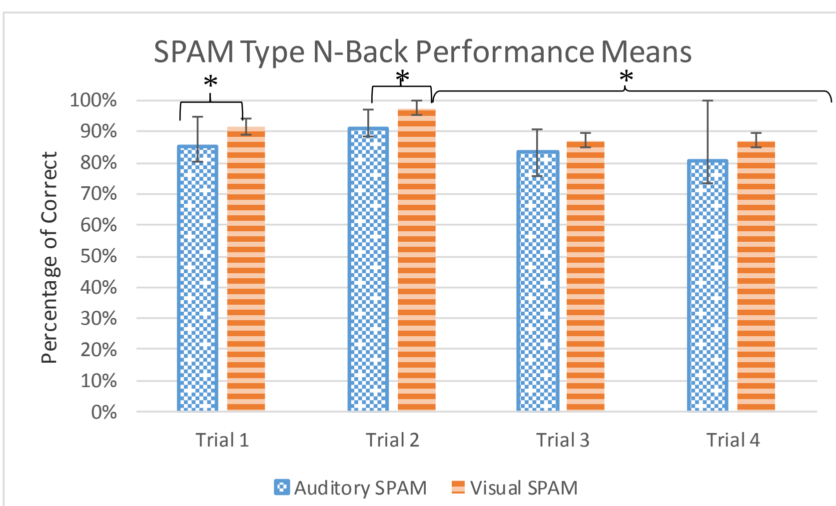

Figure 6. SPAM Type N-Back Performance Means

the Auditory and Visual SPAM groups, respectively), t (39)= .771, $p > .05$.

H 1-2B Results. There was also an expected significant decrease in performance in all measures except the N-back task when the Visual SPAM group enters into Trial 2 and Trial 3. The results indicated that presentation method of SPAM affected N-back performance (See Figure 6). The Visual SPAM group performed better than the Auditory SPAM group (mean percentage correct = 90.54 and 86.00% for Visual and Auditory SPAM group N-back performance, respectively), $F(1,38) = 15.84$, $p < 0.01$. Additionally, the more visually demanding tasks of BD and Visual SPAM caused that group to perceive their workload to be significantly higher than the Auditory SPAM group (NASA TLX workload percentage rating means = 72.16 and 76.74% for the Auditory and Visual SPAM groups, respectively; where, lower is better and indicates less perceived workload), $F(1,38) = 72.13$, $p < 0.01$.

H 2-1. The SPAM task was expected to suffer the most degradation in Trial 4 tasks in the Auditory SPAM group. In all cases, there should be significant evidence of a decrease in performance of a task when that task takes place simultaneously with another task in a like modality. The SITREP task was expected to suffer the most degradation in the Trial 4 tasks in the Auditory SPAM group. Unfortunately, due to an error in the output data coded from the SPAM interface, the results of the SPAM responses were not

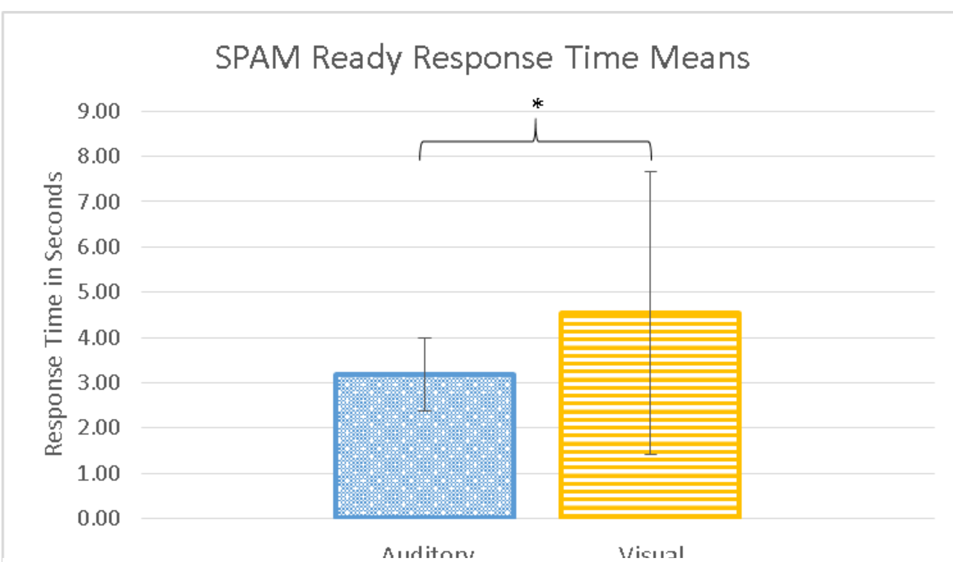

Figure 7. Between Subjects SPAM Ready Response Time Means

captured. After reviewing the data output and Durso et al., 1995, the Ready Response Time and the total SPAM Response Times were used as indicators of Situation Awareness and Workload, respectively.

Figure 7 shows the results of the N-back effects on SPAM Ready Response Times between the two groups, the Visual SPAM group reported slower Ready Response Times when compared to the Auditory SPAM group (Ready Response Time means = 3.18s and 4.54s with SD = .81s and 3.12s, for the Auditory and Visual SPAM groups, respectively), ($F(1,38)=4.855$, $p<0.05$). Figure 8 shows that the slower "Ready" state was coupled with a higher perceived workload in the Visual SPAM group (NASA TLX mean ratings = 72.16 and 76.74% with SD = 10.11 and 9.32% for the Auditory and Visual SPAM conditions), $F(1,38)=9.673$, $p<0.01$. When analyzing the results of the BD effects on SPAM response times between the two groups, the Visual SPAM group lagged far behind the Auditory SPAM Response Times indicating that the extra visual and psychomotor WL induced by the BD task increased the amount of time it took the Visual

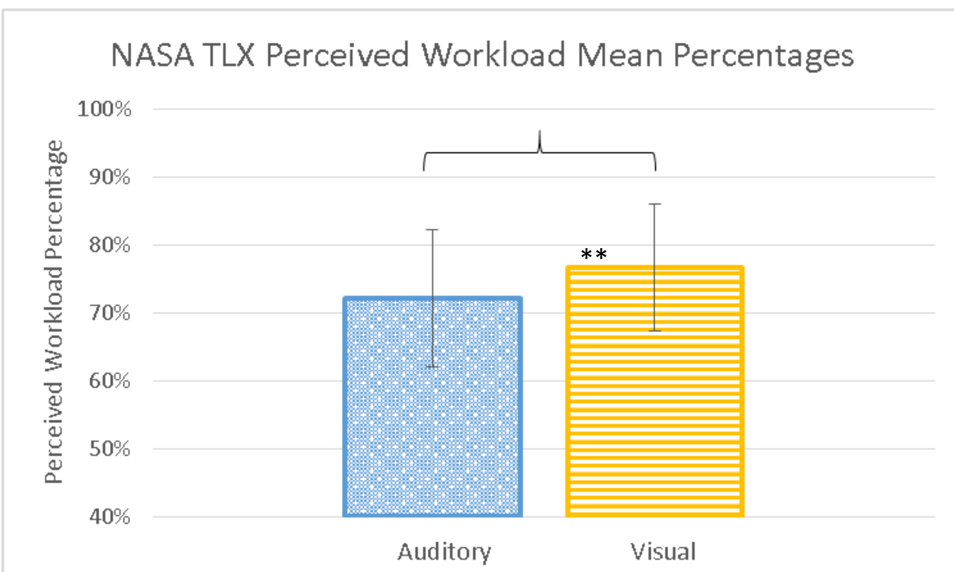

Figure 8. NASA TLX Perceived Workload Means

SPAM group to perform their SPAM tasks (SPAM Response Time means = 6.76s and 11.49s with SD = 2.35s and 2.93s, for the Auditory and Visual SPAM groups, respectively), ($F(1, 38)=4.94$, $p<0.05$).

H 2-1A. Hypothesis 2 – 1A is quite complex and must be broken down by first looking at the N-Back components of the hypothesis, followed by the BD components. In Trial 4 the Visual SPAM group should result in lower BD and N-back performance when compared to previous trials (BD more so than N-back). When looking at Figure 6, across all trials, The Visual SPAM group performed better on the N-back task than did the Auditory SPAM group (mean percentage correct = 90.54 and 86.00% for visual and auditory N-back performance, respectively), $F(1,38)=26.445$ $p<0.01$. The right most bar pair on Figure 6 shows that in Trial 4, the Auditory SPAM group might have underperformed the Visual SPAM group (Mean N-back Performance = 80.80 and 87.26 with SD = 19.04 and 7.41, respectively), but that visual difference in means was not supported by the analysis indicating that neither group actually outperformed the other ($t(40) = -1.42$, $p > .05$).

The next component of this hypothesis involves comparing the mean differences of each trial against Trial 4 N-back Performance. Figure 6 highlights that there was significant performance degradation for the Visual SPAM N-back as indicated by the significance bar over the Visual means in Trial 1 and Trial 4 (Mean N-back Performance = 91.63 and 87.26% with SD = 4.92 and 7.41%, for Trial 1 and Trial 4, respectively), t (20) = X.XX, $p < 0.01$. Figure 6 also shows that there was significant performance degradation for the Visual SPAM N-back as indicated by the significance bracket over the Visual means in Trial 2 and Trial 4 (Mean N-back Performance = 97.53 and 87.26%

with SD = 2.80 and 7.41%, respectively), t (20) = X.XX, $p < 0.01$.When comparing the Trial 3 and Trial 4 performance means in Figure 6 (the two right most Visual SPAM mean bars), there does not seem to be any evidence to support the hypothesis that the Visual SPAM group declined in performance from Trial 3 to Trial 4 (Mean N-back Performance = 87.19 and 87.26% with SD = 8.27 and 7.41%, respectively) t (20) = X.XX, $p > 0.05$.

The final component to hypothesis 2 – 1A entails a comparative analysis of BD performance of the Visual SPAM group. A quick reference to the significance bracket in Figure 5 shows that the BD performance from Trial 3 to Trial 4 did show a very large reduction in performance as expected (Mean BD performance percentages = 70.45 and 49.41 with SD = 7.02 and 11.70, in Trial 3 and 4, respectively), t (20) = X.XX, $p < 0.01$. There was no difference in means when comparing BD Performance of the Visual SPAM group's Trial 2 to Trial 4 performance (Mean BD performance percentages = 48.89 and 49.41 with SD = 8.84 and 11.70, in Trial 2 and 4, respectively), t (20) = X.XX, $p > 0.05$. The absence of change in BD performance in this group's Trial 2 and 4 confirms that N-back did not affect the BD manipulation, $F(1,38) = 2.63$, $p > .05$.

H 2-1B. In Trial 4 for the Auditory SPAM should result in lower BD and N-back performance (N-back more so than BD) when compared to previous trials and should show better performance than the Visual SPAM group. As shown (without significance brackets to reduce clutter) in Figure 5, all Trials indicated the Auditory SPAM group performed better than Visual SPAM group on the BD task, $F(1,38) = 10.38$ with a $p < .01$. Concerning the Auditory SPAM group's BD performance comparisons, the significance brackets highlight that there was a difference between Trial 1 and Trial 4

(BD performance means = 79.75 and 52.05% with SD = 7.64 and 10.21%), t (20) = X.XX, $p < 0.01$, as well as, Trial 3 and Trial 4 (BD performance means = 76.47 and 52.05% with SD = 10.34 and 10.21%), t (20) = X.XX, $p < 0.01$. The mean of 52.05 was also the lowest in this group for all Trials (See Table 1 in Appendix A). In the right most bar pair, specifically the Auditory bar, there is an extreme standard deviation in the Auditory SPAM group's N-back performance (19.04%) in Trial 4. Presumably, due to this extreme standard deviation within our participant sample, there is no evidence that the N-back performance was lower after Trial 4 than in any previous trial (For mean comparisons, See Appendix A, Table 1), t (20) = X.XX, $p > 0.05$.

CHAPTER 7

DISCUSSION

This research indicates that there are many nuances to MRT, channel loading, and Situation Awareness maintenance. One thing this research has answered, is that there is a threshold for performance and situation awareness maintenance in complex workload environments, such as those a military commander currently finds themselves. For example, the SPAM Ready Response Times indicate that the participants in the Visual group were able to respond to the visual stimuli faster due to the decreased load on their visual system when compared to the Auditory group; however, in overall length of response time was greater due to the added psycho-motor tasks of answering the SPAM question. This highlights the assertion that taxing the visual resource channels impacts the workload; while, taxing the auditory resource channels impacts the situation awareness, but less so the overall ability to handle workload due to the taxation on executive functions. Ultimately, executive function plays a more important role in determining situation awareness maintenance and task performance than does the conditioned VACP workload level.

The future will only lead to implementation of more technology and subsequent VACP workload requirements. In exploring the effects of varied workload combinations and their direct impact on performance and situation awareness, this research found that a balance must be achieved in human-centered systems. The disparity in balancing technology implements and their inherent additions to workload is something that must be addressed in order to prevent fatal decisions. The results exhibited many parallels to what was expected and supports that in designing new technology, the design must be in concert with existing systems and those systems must combine the use of auditory cues with digital interface reporting to provide a less redundant method of communication both internally and externally. Further application of these results should be used in communities exploring in-vehicle technologies and corresponding methods of auditory and visual stimuli on performance, such as in automated driver scenarios.

Future research should include increasing the level of cognitive load complexity by examining visual alerts with verbal responses and auditory alerts with psychomotor (typed) responses. Further exploration of the wealth of data from this study can answer more questions about the impact and potential correlation of subjective workload perceptions on performance and situation awareness maintenance. Future research should include an examination of VACP workload thresholds and optimal performance of individuals, as a balance must be struck in the use of technology on the battlefield. This study has shown that there is a ceiling of operational limits and that without cognizant design of human-centered systems, the human system will fail. Ideally, the results of this study will prompt discussions and further research in the military community on instituting vehicle commander battle tracking and communication visual interfaces.

APPENDIX A

Table 1. Paired t-test of between subjects means at each Trial level

Measure	Mean	Standard Deviation	N
Ball Drop Performance			
Trial 1			
Auditory	79.75	7.64	22
Visual	72.54	10.62	20
Trial 2			
Auditory	58.11	13.43	22
Visual	48.89	8.84	20
Trial 3			
Auditory	76.47	10.34	22
Visual	70.45	7.019	20
Trial 4			
Auditory	52.05	10.21	22
Visual	49.41	11.70	20
N-Back Performance			
Trial 1			
Auditory	85.11	9.60	22
Visual	91.63	4.92	20
Trial 2			
Auditory	91.05	6.01	22
Visual	97.53	2.80	20
Trial 3			
Auditory	83.71	6.83	22
Visual	87.19	8.27	20
Trial 4			
Auditory	80.80	19.04	22
Visual	87.26	7.41	20

Notes. Trial 1 = BD performance set to 90% with 0-back presentation; Trial 2= BD Performance set to 60% at 0-back; Trial 3= BD performance set to 90% at 1-back; Trial 4= BD performance set to 60% with 1-back. * $p < .05$.

Table 2. SPAM Questions presented to participants

1	What percent ammo do you have left?
2	What is your current threat status?
3	What is your current personnel status?
4	What status is currently the most dangerous?
5	Is your current location closer to the starting or end point?
6	What is your obstruction status?
7	How many targets can you engage with your remaining ammo if each needs 5%?
8	If you can travel 1 hour on 10% of fuel, how long can you travel on the current amount you have?
9	How many more attacks can you endure, if each drains you of 25% of vehicle status?

REFERENCES

Bayes, J., & Tolva, J. (1998). Sonopticon: Auditory Display in Automobile Interface Design.

Blanco, M., Biever, W. J., Gallagher, J. P., & Dingus, T. A. (2006). The impact of secondary task cognitive processing demand on driving performance. Accident Analysis and Prevention, 38, 895-906.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. Accident Analysis and Prevention, 40, 1282-1293.

Chen, J. Y., & Nayeem, R. V. (2010). Head-up display for current force vehicles (ARL-TR-5167). Adelphi, MD: Army Research Laboratory.

Cooke, Nancy J. Gorman, Jamie C. Rowe, Leah J. (2004). An Ecological Perspective on Team Cognition. COGNITIVE ENGINEERING RESEARCH INST MESA AZ.

Davies, T., & Beeharee, A. (2012). The case of the missed icon: change blindness on mobile devices.

Damos, D. (Ed.). (1991). Multiple task performance. CRC Press.

Department of Defense: 2014 Quadrennial Defense Review. (2014). Retrieved from Department Of Defense website: http://www.defense.gov/pubs/2014_Quadrennial_Defense_ Review.pdf

Durlach, P. (2003). Army digital systems and vulnerability to change blindness. Proceedings

from the 24th Army Science Conference, 1-9.

Durso, F.T., Truitt, T.R., Hackworth, C.A., Crutchfield, J.M., Nikolic, D., Moertl, P.M., Ohrt, D. & Manning, C.A. (1995). Expertise and Chess: a Pilot Study Comparing Situation Awareness Methodologies. In: D.J. Garland & M. Endsley (Ends), Experimental Analysis and Measurement of Situation Awareness. Embry-Riddle Aeronautical University Press.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., & Nikolic, D. (1999). Situation

Awareness As a Predictor of Performance in En Route Air Traffic Controllers (No.

DOT/FAA/AM-99/3). OKLAHOMA UNIV NORMAN DEPT OF PSYCHOLOGY.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37(1), 32-64. doi:10.1518/001872095779049543

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. Transportation Research Part F-traffic Psychology and Behavior, 8, 97-120. doi:10.1016/j.trf.2005.04.012

Gable, T. M., & Walker, B. N. (2013). Driving Simulator Sickness Screening: Efficient and Effective New Protocol. AutomotiveUI, Adjunct Proceedings, 13, 101-102.

Gable, T.M., Walker, B.N., Moses, H.R., & Chitloor, R.D., (2013, October). Advanced Auditory Cues on Mobile Phones Help Keep Drivers' Eyes on the Road. Proceeding of the Fifth International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Eindhoven, The Netherlands.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology,52, 139-183.

Headquarters, Department of The Army. (2012). The Operations Process (ADP 5-0). Retrieved from http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/adp5_0.pdf

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2013). Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. Frontiers in Human Neuroscience, 7, 935.

Horrey, W. J., & Wickens, C. D. (2006). Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques. Human Factors, 48(1), 196-205.

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behavior: Effects of visual and cognitive load in simulated rural driving. Transportation Research Part F-traffic Psychology and Behaviour, 8, 79-96.

Jeon, M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2009). Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009).

Jeon, M., & Walker, B. N. (2009). "Spindex": Accelerated Initial Speech Sounds Improve Navigation Performance in Auditory Menus. Human Factors and Ergonomics Society Annual Meeting Proceedings.

Jeon, M., Walker, B. N., & Gable, T. M. (2015). The effects of social interactions with in-vehicle agents on a driver's anger level, driving performance, situation awareness, and perceived workload. Applied Ergonomics, 50, 185-199.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. Multiple-task performance, 279-328.

Kramer, G., Santa Fe Institute (Santa Fe, & N.M.), I. (1994). Auditory display: Sonification, audification, and auditory interfaces. Reading, MA: Addison-Wesley.

Landsdown, T.C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. Ergonomics, 47(1), 91-104.

Liao, J., & Moray, N. (1993). A simulation study of human performance deterioration and mental workload. Le Travail Humain, 321-344.

Lim, J., Wu, W. C., Wang, J., Detre, J. A., Dinges, D. F., & Rao, H. (2010). Imaging brain fatigue from sustained mental workload: an ASL perfusion study of the time-on-task effect. Neuroimage, 49(4), 3426-3435.

Lee, Y., Lee, J. D., & Boyle, L. N. (2009). The Interaction of Cognitive Load and Attention-Directing Cues in Driving. Human Factors, 51, 271-280.

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-Based Interaction with In-Vehicle Computers: The Effect of Speech-Based E-Mail on Drivers' Attention to the Roadway. Human Factors, 43(4), 631-640.

Lenne, M. G., Hoggan, B. L., Fidock, J., Stuart, G., & Aidman, E. (2014). The Impact of Auditory Task Complexity on Primary Task Performance in Military Land Vehicle Crew. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58(1), 185-191.

Mitchell, D. K. (2000). Mental workload and ARL workload modeling tools (No. ARL-TN-161). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD.

Reyes, M. L. (2010). Examining driver performance in response to work zone interventions in a driving simulator (Master's thesis). Retrieved from http://ir.uiowa.edu/etd/729.

Robert, G., & Hockey, J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. Biological psychology, 45(1), 73-93.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. Teleoperators and Virtual Environments, 11(4), 335-351.

Santos, J., Merat, N., Mouta, S., Brookhuis, K., & Waard, D. D. (2005). The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. Transportation Research Part F-traffic Psychology and Behaviour, 8, 135-146.

Scribner, D. R., Wiley, P. H., Harper, W. H., & Kelley, T. D. (2007). The effects of workload presented via visual and auditory displays on soldier shooting and secondary task performance (No. ARL-TR-4224). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell Phone-Induced Failures of Visual Attention During Simulated Driving. Journal of Experimental Psychology-applied, 9(1), 23-32.

Sweeney, M. M. (2008). Blue Force Tracking: Building a Joint Capability. Ft. Belvoir: Defense Technical Information Center.

United States. Department of Defense. (2013). Unmanned systems integrated roadmap FY2013-2038 (14-S-0553). Retrieved from Department of Defense website: http://www.defense.gov/pubs/DOD-USRM-2013.pdf

Vu, K. P. L., & Chiappe, D. (2015). Situation awareness in human systems integration. Vu, Kim-

Phuong L.; Chiappe, Dan

Wickens, C.D. (1991). Processing resources and attention. In D.L. Damos (Ed.), Multiple Task Performance, 3-34. Washington DC: Taylor & Francis.

Wickens, C.D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomic Sciences 3 (2), 159-177.

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 397-403.

Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 449-455.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). Mental workload, stress, and individual differences: cognitive and neuroergonomic perspectives. In Engineering Psychology and human performance (4th ed., p. 352). Psychology Press.

Wickens, C. D., & Huey, B. M. (Eds.). (1993). Workload Transition: Implications for Individual and Team Performance. National Academies Press.

Zhang, Zhen, Walji, Muhummad F, P., Gimbel, R., & Zhang, J. (Eds.). (n.d.). Functional Analysis of Interfaces in U.S. Military Electronic Health Record System using UFuRT Framework. American Medical Informatics Association.